Package 'HDMT'

February 20, 2021

Type Package

Title A Multiple Testing Procedure for High-Dimensional Mediation Hypotheses

Version 1.0.3

Date 2021-02-20

Author James Dai [aut, cre],
Xiaoyu Wang [aut]

Maintainer James Dai <jdai@fredhutch.org>

Description A multiple-testing procedure for high-dimensional mediation hypotheses. Mediation analysis is of rising interest in epidemiology and clinical trials. Among existing methods for mediation analyses, the popular joint significance (JS) test yields an overly conservative type I error rate and therefore low power. In the R package 'HDMT' we implement a multiple-testing procedure that accurately controls the family-wise error rate (FWER) and the false discovery rate (FDR) when using JS for testing high-dimensional mediation hypotheses. The core of our procedure is based on estimating the proportions of three component null hypotheses and deriving the corresponding mixture distribution of null p-values. Results of the data examples include better-behaved quantile-quantile plots and improved detection of novel mediation relationships on the role of DNA methylation in genetic regulation of gene expression. With increasing interest in mediation by molecular intermediaries such as gene expression, the proposed method addresses an unmet methodological challenge. Methods used in the package refer to James Y. Dai, Janet L. Stanford & Michael LeBlanc (2020) <doi:10.1080/01621459.2020.1765785>.

Depends R (>= 3.4.0)

Imports fdrtool

LazyLoad no

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2021-02-20 22:30:02 UTC

R topics documented:

`adjust_quantile` *A function to compute the quantiles of the estimated mixture null distribution for pmax using either the approximation or exact method*

Description

A function to generate the quantiles of the estimated three-component mixture null distribution for pmax (the maximum of the two p-values for testing mediation) using either approximation or exact method

Usage

```
adjust_quantile(alpha00, alpha01, alpha10, alpha1, alpha2,
input_pvalues, exact = 0)
```

Arguments

`alpha00`	A numeric number represents the proportion of null H_{00}
`alpha01`	A numeric number represents the proportion of null H_{01}
`alpha10`	A numeric number represents the proportion of null H_{10}
`alpha1`	A numeric number represents the proportion of null alpha==0 (association between exposure and mediator)
`alpha2`	A numeric number represents the proportion of null beta==0 (association between mediator and outcome adjusted for exposure)
`input_pvalues`	A matrix contains two columns of p-values for candidate mediators. Column 1 is the p-value of testing if a exposure is associated with the mediator (alpha!=0). Column 2 is the p-value of testing if a mediator is associated with the outcome adjusted for exposure(beta!=0)
`exact`	Use the option to choose from two methods. exact=0: the approximation method without estimating the CDFs when deriving the mixture null distribution; exact=1: the exact method to estimate the CDFs nonparametrically when deriving the mixture null distribution

Details

The is a function to computed the expected quantiles for the observed p-max values based on the estimated mixture null distribution. The methodology detail can be found in Dai et al (2020).

Value

A vector contains the expected quantiles of p-values based on the estimated mixture null distribution. See Dai et al (2020) for details of how to compute quantiles using the approximation method (exact=0) or the exact method (exact=1).

Author(s)

James Y. Dai and X. Wang

References

James Y. Dai, Janet L. Stanford, Michael LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses. Journal of the American Statistical Association. 2020, In Press.

Examples

```
data(snp_input)
input_pvalues <- snp_input

#To save time for illustration, we use 10 percent of rows

input_pvalues <- input_pvalues[sample(1:nrow(input_pvalues),size=ceiling(nrow(input_pvalues)/10)),]

nullprop <- null_estimation(input_pvalues,lambda=0.5)

pnull <- adjust_quantile(nullprop$alpha00,nullprop$alpha01,nullprop$alpha10,nullprop$alpha1,
                         nullprop$alpha2,input_pvalues,exact=0)
```

`correct_qqplot` *A function to draw the corrected quantile-quantile plot for p-max using the expected quantiles*

Description

A function to draw the corrected quantile-quantile (Q-Q) plots. The corrected quantiles were computed from the mixture null distribution (green dots) and the naive quantiles were computed from the uniform distribution (red dots).

Usage

```
correct_qqplot(pmax, pnull, opt="all")
```

Arguments

pmax	The vector for maximum p-values
pnull	The quantiles of pmax based on the estimated mixture null distribution
opt	Option to draw the plot. opt="all":use all the data points, opt="subset": use a subset of the data points, in case there are too many points in a genome-wide setting, to avoid drawing an overcrowded Q-Q plot with a prohibitive image size.

Author(s)

James Y. Dai and X. Wang

References

James Y. Dai, Janet L. Stanford, Michael LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses, Journal of the American Statistical Association. 2020. In Press.

Examples

```
data(snp_input)
input_pvalues <- snp_input
#To save time for illustration, we use 10 percent of rows
input_pvalues <- input_pvalues[sample(1:nrow(input_pvalues),
                                size=ceiling(nrow(input_pvalues)/10)),]

pmax <- apply(input_pvalues,1,max)
nullprop <- null_estimation(input_pvalues,lambda=0.5)
pnull1 <- adjust_quantile(nullprop$alpha10,nullprop$alpha01,nullprop$alpha00,
                          nullprop$alpha1,nullprop$alpha2,input_pvalues,exact=1)
correct_qqplot(pmax,pnull1)
```

---

exercise_input *An example dataset to demonstrate the usage of 'HDMT'*

Description

This example dataset was included to assess the mediation role of DNA methylation in the effect of exercise on prostate cancer progression in a Seattle-based cohort of patients diagnosed with clinically localized PCa. The entire data set contains two sets of p-values from genome-wide testing of 450K CpG sites. Due to space limit, a subset (10 percent) of the full dataset is included in the package for illustration.

The dataset is a matrix containing two columns of p-values for candidate mediators. Column 1 contains the p-values for testing if an exposure is associated with the mediator (alpha!=0). Column 2 contains the p-value for testing if a mediator is associated with the outcome after adjusted for the exposure (beta!=0).

Usage

```
data("exercise_input")
```

Format

The format of exercise_input is: num [1:47900, 1:2] 0.4966344 0.1048730 0.1005355 0.4946623 ...

References

James Y. Dai, Janet L. Stanford, Michael LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses. Journal of the American Statistical Association, 2020, In Press.

Examples

```
data(exercise_input)
dim(exercise_input)
```

---

fdr_est *A function to compute the estimated pointwise FDR for every observed p-max*

Description

A function to compute the estimated pointwise FDR based on the proposed joint significance mixture null method (JS-mixture).

Usage

```
fdr_est(alpha00, alpha01, alpha10, alpha1, alpha2, input_pvalues, exact = 0)
```

Arguments

alpha00	A numeric number represents the proportion of null H_{00}
alpha01	A numeric number represents the proportion of null H_{01}
alpha10	A numeric number represents the proportion of null H_{10}
alpha1	A numeric number represents the proportion of null alpha=0
alpha2	A numeric number represents the proportion of null beta=0
input_pvalues	A matrix contains two columns of p-values for candidate mediators. Column 1 is the p-value of testing if an exposure is associated with the mediator (alpha!=0). Column 2 is the p-value of testing if a mediator is associated with the outcome adjusted for the exposure (beta!=0)
exact	The option to choose from two methods. exact=0: approximation without estimating the CDFs; exact=1: estimate the CDFs nonparametrically

Details

A function to estimate the pointwise FDR based on the proposed method to estimate the mixture null distribution. See Dai et al (2020) for details of how to compute quantiles using the approximation method (exact=0) or the exact method (exact=1).

Value

The estimated pointwise FDR for p-max

Author(s)

James Y. Dai and X. Wang

References

James Y. Dai, Janet L. Stanford, Michael LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses, Journal of the American Statistical Association. 2020. In press.

Examples

```
data(snp_input)
input_pvalues <- snp_input
#To save time for illustration, we use 10 percent of rows
input_pvalues <- input_pvalues[sample(1:nrow(input_pvalues),size=ceiling(nrow(input_pvalues)/10)),]

nullprop <- null_estimation(input_pvalues,lambda=0.5)
fdr <- fdr_est(nullprop$alpha00,nullprop$alpha01,nullprop$alpha10,
               nullprop$alpha1,nullprop$alpha2,input_pvalues,exact=0)
```

---

fwer_est *A function used to compute Family wise error rate (FWER) cutoff for p-max at a designated level*

Description

A function to compute the FWER cutoff for p-max using the estimated mixture null distribution

Usage

```
fwer_est(alpha10, alpha01, alpha00, alpha1, alpha2, input_pvalues,
         alpha = 0.05, exact = 0)
```

Arguments

alpha00	A numeric number represents the proportion of null H_{00}
alpha01	A numeric number represents the proportion of null H_{01}
alpha10	A numeric number represents the proportion of null H_{10}
alpha1	A numeric number represents the proportion of null alpha==0
alpha2	A numeric number represents the proportion of null beta==0

input_pvalues	A matrix contains two columns of p-values for candidate mediators. Column 1 is the p-value of testing if an exposure is associated with the mediator (alpha!=0). Column 2 is the p-value of testing if a mediator is associated with the outcome adjusted for the exposure (beta!=0)
alpha	The designated significance level for FWER
exact	The option to choose from two methods. exact=0: approximation without estimating the CDFs; exact=1: estimate the CDFs nonparametrically

Details

A function to compute FWER cutoff for p-max accounting for the mixture null distribution. The methodology detail can be found in Dai et al (2020).

Value

A numeric number represents the output FWER cutoff

Author(s)

James Y. Dai and X. Wang

References

James Y. Dai, Janet L. Stanford, Michael LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses, Journal of the American Statistical Association. 2020. In Press.

Examples

```
data(snp_input)
input_pvalues <- snp_input
#To save time for illustration, we use 10 percent of rows
input_pvalues <- input_pvalues[sample(1:nrow(input_pvalues),
                               size = ceiling(nrow(input_pvalues)/10)),]

nullprop <- null_estimation(input_pvalues,lambda=0.5)
fwercut0 <- fwer_est(nullprop$alpha10,nullprop$alpha01,nullprop$alpha00,nullprop$alpha1,
nullprop$alpha2,input_pvalues,alpha=0.05,exact=0)

fwercut1 <- fwer_est(nullprop$alpha10,nullprop$alpha01,nullprop$alpha00,nullprop$alpha1,
nullprop$alpha2,input_pvalues,alpha=0.05,exact=1)
```

`null_estimation` *A function to estimate the proportions of the three component nulls*

Description

This is a function to estimate the proportions of the three component nulls involved in mediation testing. We developed a three component-mixture model method to estimate the proportions of nulls and provide much more accurate control of the family-wise error rate (FWER) and the false discovery rate (FDR), when compared to the standard approach using the uniform null distribution.

Usage

```
null_estimation(input_pvalues, lambda = 0.5)
```

Arguments

`input_pvalues`	A matrix contains two columns of p-values for candidate mediators. Column 1 is the p-value of testing if the exposure is associated with the candidate mediator (alpha!=0). Column 2 is the p-value of testing if the candidate mediator is associated with the outcome adjusted for the exposure (beta!=0).
`lambda`	A tuning parameter between 0 and 1, the default value is 0.5.

Details

A function to estimate the proportions of the three types of component null hypotheses:

H_{00}: alpha==0 and beta==0

H_{01}: alpha==0 and beta!=0

H_{10}: alpha!=0 and beta==0

The methodology detail can be found in Dai et al (2020).

Value

A list contains five elements.

`alpha00`	A numeric number represents the proportion of null H_{00}
`alpha01`	A numeric number represents the proportion of null H_{01}
`alpha10`	A numeric number represents the proportion of null H_{10}
`alpha1`	A numeric number represents the proportion of null alpha==0
`alpha2`	A numeric number represents the proportion of null beta==0

Author(s)

James Y. Dai and X. Wang

References

James Y. Dai, Janet L. Stanford, Michael LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses. Journal of the American Statistical Association, 2020, In Press.

Examples

```
data(snp_input)
input_pvalues <- snp_input
#To save computing time for illustration, we use 10 percent of rows (p-values)
input_pvalues <- input_pvalues[sample(1:nrow(input_pvalues),
size <- ceiling(nrow(input_pvalues)/10)),]

nullprop <- null_estimation(input_pvalues,lambda=0.5)
```

---

snp_input *An example dataset to demonstrate the usage of 'HDMT'*

Description

This example dataset is included in 'HDMT' to assess the mediation role of DNA methylation in genetic regulation of gene expression in primary prostate cancer (PCa) samples from The Cancer Genome Atlas (TCGA) with risk SNPs as the exposure.

The dataset is a matrix containing two columns of p-values for candidate mediators. Column 1 contains the p-values for testing if an exposure is associated with the mediator (alpha!=0). Column 2 contains the p-value for testing if a mediator is associated with the outcome after adjusted for the exposure (beta!=0).

Usage

```
data("snp_input")
```

Format

The format of snp_input is: num [1:69602, 1:2] 0.106 0.999 0.101 0.173 0.89 ...

References

James Y. Dai, Janet L. Stanford, Michael LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses. Journal of the American Statistical Association, 2020, In Press.

Examples

```
data(snp_input)
dim(snp_input)
```

Index